Kohonen's Map Approach for the Belief Mass Modeling

Imen Hammami, Grégoire Mercier, Atef Hamouda, Jean Dezert

Kohonen's Map Approach for the Belief Mass Modeling

Imen Hammami, Grégoire Mercier, *Senior Member, IEEE*, Atef Hamouda and Jean Dezert

***Abstract*—In the framework of the evidence theory, several approaches for estimating belief functions are proposed. However, they generally suffer from the problem of masses attribution in case of compound hypotheses that lose much conceptual contribution of the theory. In this paper, an original method for estimating mass functions using Kohonen's map derived from the initial feature space and an initial classifier is proposed. Our approach allows a smart mass belief assignment, not only for simple hypotheses, but also for disjunctions and conjunctions of hypotheses. Thus, it can model at the same time ignorance, imprecision and paradox. The proposed method for basic belief assignment (BBA) is of interest for solving estimation mass functions problems where a large quantity of multi-variate data is available. Indeed, the use of Kohonen map simplifies the process of assigning mass functions. The proposed method has been compared to state-of-the art BBA technique on benchmark database and applied on remote sensing data for image classification purpose. Experimentation shows that our approach gives similar or better results than other methods presented in the literature so far, with an ability to handle large amount of data.**

***Index Terms*—Evidence theory, Mass belief assignment, Kohonen map, Estimation.**

I. Introduction

In several fields, one observes the presence of multiple information, coming from various sources, which represents the same scene. In order to synthesize more useful information related to the phenomenon observed, it is often necessary to exploit the redundancy and complementarity of the sources. Many formalisms were proposed manipulating those sources and allowing to formalize mathematically uncertain and imprecise data as the Bayesian theory, fuzzy set theory,...

The belief function theory, introduced by Dempster [1] and formalized by Shafer [2], presents an appealing mathematical background in information fusion domain. It allows the processing of both imprecise and uncertain information stemming from very varied sources. The initial theory was modified and ameliorated on several occasions, for example through the work of Dezert-Smarandache [3], a paradoxical reasoning is added. Despite, the fact that belief function theory excels in extracting the most truthful proposition from a multisource context, the estimation of basic belief assignments has always been a difficulty for applying efficiently belief functions in applications.

In belief function estimation, we distinguish two main family approaches. Likelihood based approaches [2], [4], require the knowledge, or the estimation, of the conditional probability density for each class. The second family is the distance based approaches [5], [6]. However, these two types of estimation present some limits: among them we can mention the need of the *a priori* knowledge on the hypotheses which is not always easy to know, especially, for compound hypotheses.

In this work, we propose to define a new approach for estimating mass functions in the case of representing knowledge in complex systems, where quantity of information is important (*i.e.* a complex feature space $\mathbb{R}^p$). The construction of mass function can be done through Kohonen's Map [7] that allows to approximate the feature space dimension into a projected 2D space (so called map). Thus, the use of Kohonen's map simplifies the process of assigning mass functions on conjunctions and disjunction of hypotheses when considering relative distance of an observation to the map. In the feature space (in $\mathbb{R}^p$), operations on basic belief assignment (BBA) can be much more complex and may not be feasible due to computing time or accuracy consideration.

This paper is organized as follows. The second section briefly introduces the main concepts of the belief function theory. We survey some existing methods for estimating mass functions in section II-B. Then section III introduces the main ideas of the proposed approach and explains the underlying methodology. Section IV provides simple examples to illustrate the methodology. In section V, the results obtained by the proposed approach are compared to some state-of-the art methods on a set of benchmark database. Then, section VI presents a deeper analysis of the classification results on a large SPOT image. Finally, section VII concludes.

II. Evidence theory

The belief function theory or the evidence theory was introduced by Dempster [1] in order to represent some imprecise probabilities with upper and lower probabilities. Then, it was mathematically formalized by Shafer [2] thanks to the general formalism of belief functions. In this section, the basic mathematical elements of this theory are presented as well as some existing methods for estimating mass functions.

A. Basic concepts

Dempster-Shafer theory (DST) is used for representing belief on imperfect observation (such as uncertain, imprecise

I. Hammami and G Mercier are with Telecom Bretagne, Brest 29238, France (e-mail: imen.hammammi@gmail.com; gregoire.mercier@telecom-bretagne.eu).

I. Hammami and A. Hamouda are with University of Tunis El Manar, Faculty of Sciences of Tunis, LIPAH , 2092, Tunis , Tunisie (e-mail: imen.hammammi@gmail.com; atef.hamouda@fst.rnu.tn).

J. Dezert is with ONERA - The French Aerospace Lab, Palaiseau 91761, France (e-mail: jean.dezert@onera.fr).

and/or incomplete) through the basic belief assignment (BBA, also called mass function), defined on all the subsets of the frame of discernment Θ, noted 2^Θ. In our context, $m(\cdot)$ will have to be built from the observation provided by a sensor, that is from a sample $\boldsymbol{x} \in \mathbb{R}^p$. A BBA $m(\cdot)$ is the mapping from elements of the power set 2^Θ onto $[0,1]$ such that:

$$m : 2^\Theta \rightarrow [0,1]$$

under constraints:

$$\begin{cases} m(\boldsymbol{x} \in \emptyset) = 0 \\ \sum_{A \subseteq 2^\Theta} m(\boldsymbol{x} \in A) = 1. \end{cases} \tag{1}$$

In this section, we use the notation $m(A)$ that stands for $m(\boldsymbol{x} \in A)$ when there is no ambiguity. The frame of discernment Θ is the set of possible answers of the problem under concern. It is composed of exhaustive and exclusive hypotheses: $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$ corresponding to Shafer's model of the frame. From this frame of discernment, a power set noted 2^Θ can be built, including all the disjunctions of hypotheses θ_i such as $\theta_i \cup \theta_j$ or $\theta_i \cup \theta_j \cup \theta_k$...

Let now consider two sources of information through their mass function m_1 and m_2 defined on the same frame of discernment. These BBAs can be combined using the conjunctive rule defined by:

$$\text{Disjunctive rule:} \quad m_{1 \bigcirc\!\!\!\!\cup 2}(A) = \sum_{\substack{E,F \in 2^\Theta \\ E \cup F = A}} m_1(E)\, m_2(F), \tag{2}$$

$$\text{Conjunctive rule:} \quad m_{1 \bigcirc\!\!\!\!\cap 2}(A) = \sum_{\substack{E,F \in 2^\Theta \\ E \cap F = A}} m_1(E)\, m_2(F). \tag{3}$$

The mass $K = \sum_{E \cap F = \emptyset} m_1(E)\, m_2(F)$ is called the degree of conflict between m_1 and m_2. Various works deal the problem of the appearance of conflict with this combination.

In the framework of Dezert-Smarandache theory (DSmT) [3] the exclusivity constraints imposed upon the hypotheses, and the redistribution of the conflicting mass to the non-empty sets using the normalization rule has been canceled. The main idea of DSmT is to work on the hyper-powerset of the frame of discernment. The hyper-power set D^Θ is defined as the set of all composite possibilities built from Θ which $\cap$ and $\cup$ operators such that:

1) $\emptyset, \theta_1, \ldots, \theta_N \in D$.
2) $\forall E \in D^\Theta, F \in D^\Theta, (E \cup F) \in D^\Theta, (E \cap F) \in D^\Theta$.
3) No other elements belong to D^Θ, except those, obtained by using rules 1 or 2.

As in DST, different combination rules were proposed in DSmT. Interested readers could refer to [8], [9] for more details about some of these rules. For decision making from combined mass function, the Generalized Pignistic Transformation [3] noted $BetP$ is used:

$$BetP(A) = \sum_{E \in D^\Theta} \frac{C_M(E \cap A)}{C_M(E)}\, m(E), \quad \forall A \in D^\Theta \tag{4}$$

where C_M is the cardinality within DSmT as defined in [10, sec. 3.2.2, p. 52]. The decision is taken by the maximum of pignistic probability function $BetP(\cdot)$. Similarly, the Pignistic Transformation [11] can be used within DST framework for decision making. Some other decision rules can be apply by using either the maximum of belief (Bel) or on the maximum of the plausibility (Pl) which are defined by:

$$Bel(A) = \sum_{\substack{E,F \in 2^\Theta \\ E \subseteq A, E \neq \emptyset}} m(E), \qquad \forall A \in 2^\Theta \tag{5}$$

$$Pl(A) = \sum_{\substack{E,F \in 2^\Theta \\ E \cap A \neq \emptyset}} m(E), \qquad \forall A \in 2^\Theta \tag{6}$$

where $Bel(A)$ represents the sum of masses of belief of the focal elements which involve A and $Pl(A)$ quantifies the maximal degree of belief that could be given to A. Eq. (6) and eq. (5) can be applied on D^Θ as well.

In the rest of this paper, decisions are given by the maximum of Pignistic Transformation in 2^Θ and by the maximum of Generalized Pignistic Transformation in D^θ.

B. Estimation of mass functions in evidence theory

Apart the choice of the fusion rule, the major difficulty of belief theory lies in estimating mass functions. Several methods have been proposed in the literature and their choice must be made depending on the nature of data and the application. As stated before, mass belief assignment can be classified in two axes: distance-based approaches and likelihood-based approaches. In this part some approaches of these categories are browsed in details.

1) Distance-based approaches: Distances-based Models correspond to models where masses relative to data depend on distances calculated in the feature space. Within the context of the theory of belief functions, three models have been introduced in the literature. One is based on the algorithm of K-Nearest Neighbor (K-NN), the other is based on the clustering method C-means, and finally the EVCLUS algorithm that assigns a BBA to each object from the matrix of dissimilarities between objects.

a) BBA with a K-NN algorithm: In this estimation approach, only the singleton θ_n and the whole frame of discernment Θ are considered. Focal elements[1] and the mass functions are estimated from a learning set $\mathcal{L} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L\}$ for which their corresponding class is known: $\boldsymbol{x}_\ell$ is assigned to class $\theta_{\boldsymbol{x}_\ell}$ among $\{\theta_1, \theta_2, \ldots, \theta_N\}$. For each instance $\boldsymbol{x}$ to be classified, the K-NN is used to retain only the closest vectors of $\boldsymbol{x}$. Let $\mathcal{N}_K(\boldsymbol{x})$ be the set of the K nearest neighbors of $\boldsymbol{x}$ in $\mathcal{L}$. This set can be considered to pieces of evidence regarding the class of $\boldsymbol{x}$. For each element $\boldsymbol{x}_k$ in $\mathcal{N}_K(\boldsymbol{x})$ ($\boldsymbol{x}_k$ being assigned to class $\theta_{\boldsymbol{x}_k}$) the strength of this evidence decreases with the distance $d(\boldsymbol{x}, \boldsymbol{x}_k)$. The BBAs are then given by the following expression:

$$\begin{cases} m_k(\boldsymbol{x} \in \theta_{\boldsymbol{x}_k}) = \alpha \varphi_{\theta_{\boldsymbol{x}_k}}(d(\boldsymbol{x}, \boldsymbol{x}_k)) \\ m_k(\boldsymbol{x} \in \Theta) = 1 - \alpha \varphi_{\theta_{\boldsymbol{x}_k}}(d(\boldsymbol{x}, \boldsymbol{x}_k)) \end{cases} \tag{7}$$

where $0 < \alpha < 1$ is a constant. $\varphi_{\theta_n}(\cdot)$ is a decreasing function verifying $\varphi_{\theta_n}(0) = 1$ and $\lim_{d \rightarrow \infty} \varphi_{\theta_n}(d) = 0$, $d(\boldsymbol{x}, \boldsymbol{x}_k)$

[1] A focal element is an element X of 2^Θ or D^Θ such that $m(X) > 0$.

being the distance between the vector $\boldsymbol{x}_k$ and $\boldsymbol{x}$. The $\varphi_{\theta_n}(\cdot)$ function might be an exponential function following this form:

$$\varphi_{\theta_n}(d) = \exp(-\gamma_n d^2) \tag{8}$$

where γ_n is a positive parameter determined separately for each class $\theta_n \in \{\theta_1, \theta_2, \ldots, \theta_N\}$. Typically in DST framework, the combined belief function $m(\cdot)$ is obtained by the application of Dempster combination operator on each sources of evidence (*i.e.* partial information) $m_k(\cdot)$.

$$m = \oplus_{k\in[1,\ldots,K]} m_k \tag{9}$$

where $\oplus$ stands for Dempster's rule defined by:

$$m_{1\oplus 2}(A) = \frac{1}{1-K} m_{1\bigcirc 2}(A) \tag{10}$$

The described method defines the Distance Classifier (DC) [5]. Despite its promising results, this approach has a major shortcoming because it cannot deal with new (exploratory) data. This point may be explained by the cost of this algorithm which is quite high because it has to calculate the Euclidean distance to each of the neighbors, and sort them to find the nearest K. This task has a computational complexity of $O(L \times p)$ for each new BBA, where p is the space dimension.

b) BBA with ECM algorithm: In [12], Denœux and Masson propose a new automatic classification method called ECM (Evidential C-Means). Let $\mathcal{L} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L\}$ be a collection of vectors in $\mathbb{R}^p$ describing the L observations. Let K be the desired number of classes. Each cluster is represented by a prototype or a center $\boldsymbol{v}_k \in \mathbb{R}^p$. Let V denotes a matrix of size $(K \times p)$ composed of the coordinates of the cluster centers such that $V_{k,q}$ is the q^{th} component of the cluster center $\boldsymbol{v}_k$. ECM looks for matrices $M = (m_{\ell,k})$ (mass functions matrix of dimension $(L \times K)$ with elements $m_{\ell,k} = m(\boldsymbol{x}_\ell \in \theta_k)$) and V by minimizing the following objective function:

$$J_{\text{ECM}}(M, V) = \sum_{\ell=1}^{L} \sum_{\substack{k=1 \\ \theta_k \subseteq \Theta, \theta_k \neq \emptyset}}^{K} c_k^\alpha m_{\ell,k}^\beta d^2(\boldsymbol{x}_\ell, \boldsymbol{v}_k) + \sum_{\ell=1}^{L} \delta^2 m_{k,\emptyset}^\beta \tag{11}$$

subject to the constraint:

$$\sum_{\substack{k=1 \\ \theta_k \subseteq \Theta, \theta_k \neq \emptyset}}^{K} m_{\ell,k} + m_{\ell,\emptyset} = 1, \quad \forall \ell \in \{1, \ldots, L\}, \tag{12}$$

where $m_{\ell,\emptyset}$ stands for $m(\boldsymbol{x}_\ell \in \emptyset)$, δ controls the amount of data considered as outliers, β is a weighting exponent that controls the imprecision of the partition and α is a parameter to control the degree of penalization. The c_k^α coefficient is a penalty factor that prevents from high cardinality class.

This algorithm holds a great importance in processing complex and imprecise data since it allows the allocation of the masses to the different subsets of the frame of discernment. Unfortunately, it has an exponential complexity relative to the number of classes and linear complexity relative to the number of samples.

c) BBA with EVCLUS algorithm: Let us consider two BBAs m_i and m_j regarding the class membership of two observations $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The aim of EVCLUS (EVidential CLUStering of proximity data) BBA estimation is: the more similar the observations, the lower the degree of conflict between their mass function and the higher plausible that they belong to the same class. As shown in [13], this idea can be explained as follows. Let R_{ij} be the following proposition "samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ belong to the same class" corresponding to the following subset of the Cartesian product $\Theta^2 = \Theta \times \Theta$:

$$R_{ij} = \{(\theta_1, \theta_1), (\theta_2, \theta_2), \ldots, (\theta_K, \theta_K)\}.$$

The plausibility $Pl_{i\times j}$ of the proposition R_{ij} can be shown to be equal to:

$$\begin{aligned} Pl_{i\times j}(R_{ij}) &= \sum_{\substack{A\times B \in \Theta^2 \\ (A\times B)\cap R_{ij} \neq \emptyset}} m_{i\times j}(A \times B) \\ &= \sum_{A\cap B \neq \emptyset} m_i(A) m_j(B) \\ &= 1 - \sum_{A\cap B = \emptyset} m_i(A) m_j(B) = 1 - \mathcal{K}_{ij} \end{aligned}$$

where $m_{i\times j}(A \times B)$ is the BBA that describes ones beliefs regarding the class membership of both samples and $\mathcal{K}_{ij}$ is the degree of conflict between m_i and m_j.

Let us assume that the available data consist of a $L \times L$ dissimilarity matrix $D = (d_{ij})$, EVCLUS looks for $M = (m_1, m_2, \ldots, m_L)$ the credal partition of $\mathcal{L} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L\}$ a set of L observations to be classified in Θ by minimizing an stress function inspired from multidimensional scaling (MDS) methods [14] such that the degree of conflict $\mathcal{K}_{ij}$ represents a form of distance between the observations and reflects the dissimilarities d_{ij}. The stress function to be minimized is given by:

$$J_{\text{EVCLUS}}(M, a, b) = \frac{1}{Ct} \sum_{i<j} \frac{(a\mathcal{K}_{ij} + b - d_{ij})^2}{d_{ij}}, \tag{13}$$

where a and b are two coefficients, d_{ij} is the dissimilarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and Ct is a constant defined for normalization as:

$$Ct = \sum_{i<j} d_{ij}.$$

Thus, EVCLUS can be thought of as an iterative optimization, with respect to M, a and b, under the criterion of eq. (13) to be minimized by using a gradient-based procedure. The major drawback of this algorithm is its computational complexity thus it is limited to data sets of a few thousand elements and less than 20 classes.

2) Likelihood-based approaches: In [15], among the several probabilistic models that have been proposed in the literature, Appriou considers each class as a particular source of information. Then the mass is defined through the transfer of the bayesian probability function to the total ignorance and the complementary class[2]. Then the BBA associated to the hypothesis θ_k is defined through the source of information S_k

[2] *i.e.* θ and θ^c are complementary if $\theta \cup \theta^c = \Theta$ and $\theta \cap \theta^c = \emptyset$.

with the following mass:

$$\begin{cases} m_k(\boldsymbol{x} \in \theta_k) = \alpha_k \frac{Rp(\boldsymbol{x} \in \theta_k)}{1+Rp(x \in \theta_k)} \\ m_k(\boldsymbol{x} \in \theta_k^c) = \alpha_k \frac{1}{1+Rp(x \in \theta_k)} \\ m_k(\boldsymbol{x} \in \Theta) = 1 - \alpha_k, \end{cases} \tag{14}$$

where α_k with values in $[0, 1]$ is a discounting factor associated with the reliability of the model to the class θ_k and $R = 1/\max_k p(\boldsymbol{x} \in \theta_k)$ is a positive normalized coefficient less or equal to 1. The classes θ_k^c are defined in 2^Θ such as $\theta_k \cap \theta_k^c = \emptyset$. From these K belief functions, each elementary sources are fused by using the orthogonal (disjunctive) sum given in eq. (9), yielding a complete BBA on 2^Θ. In [16] a transfer model is introduced to distribute the initial masses over the compound hypotheses (disjunction of classes).

III. A NEW METHOD TO BUILD BBA

As previously discussed, the belief function theory provides a robust framework for uncertain information modeling. Nevertheless, the application of this theory in fusing sources of information induces some well-known problems. One of them is the problem of estimating the belief functions. The next section gives an overview on Kohonen's Self Organizing Map (SOM)that will be used to solve this problem. Section III-B will present the feature space that is defined to help the estimation of mass functions. The BBA itself is detailed on section III-C.

A. Overview on Kohonen's map

There exist many versions of the SOM. However, the basic philosophy is very simple and already effective [7]. A SOM defines a mapping from the input feature space (say $\mathbb{R}^p$) onto a regular array of $M \times N$ nodes (see Fig. 1) [17].

A reference vector, also called weight vector, $\boldsymbol{w}(i, j) \in \mathbb{R}^p$ is associated to the node at each position (i, j) with $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant M$. An input vector $\boldsymbol{x} \in \mathbb{R}^p$ is compared to each $\boldsymbol{w}(i, j)$. The best match is defined as output of the SOM: thus,

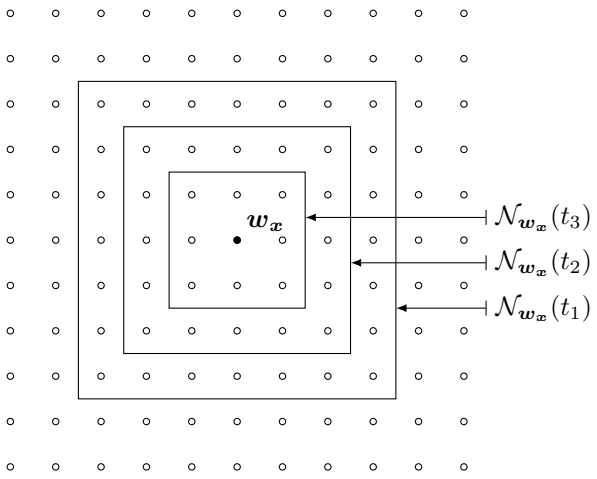


Fig. 1. A schematic view of a 11×11 Kohonen's Self Organizing Map. Several topological neighborhood $\mathcal{N}_{\boldsymbol{w}_{\boldsymbol{x}}}(t_i)$ of the winning neuron $\boldsymbol{w}_{\boldsymbol{x}}$ are drawn. The size is decreasing with the number of iterations ($t_1 < t_2 < t_3$) during the training phase, according to (17).

the input data $\boldsymbol{x}$ is mapped onto the SOM at location $(i_{\boldsymbol{x}}, j_{\boldsymbol{x}})$ where $\boldsymbol{w}(i_{\boldsymbol{x}}, j_{\boldsymbol{x}})$ is the neuron the most similar to $\boldsymbol{x}$ according to a given metric. SOM performs a non linear projection of the probability density function $p(\boldsymbol{x})$ from the high-dimensional input data onto the two-dimensional array.

In practical applications, the Euclidean distance is usually used to compare $\boldsymbol{x}$ and $\boldsymbol{w}(i, j)$ in $\mathbb{R}^p$, so that $d(\boldsymbol{x}, \boldsymbol{w}(i, j)) = \|\boldsymbol{x} - \boldsymbol{w}_{\boldsymbol{x}}\|$. The node that minimizes the distance between $\boldsymbol{x}$ and $\boldsymbol{w}(i, j)$ defines the best-matching node (or the so-called winning neuron), and is denoted by the subscript $\boldsymbol{w}_{\boldsymbol{x}}$:

$$d(\boldsymbol{x}, \boldsymbol{w}_{\boldsymbol{x}}) = \|\boldsymbol{x} - \boldsymbol{w}_{\boldsymbol{x}}\| = \min_{\substack{1 \leqslant i \leqslant M \\ 1 \leqslant j \leqslant N}} \|\boldsymbol{x} - \boldsymbol{w}(i, j)\|. \tag{15}$$

An optimal mapping would be the one that maps the probability density function $p(\boldsymbol{x})$ in the most faithful fashion, preserving at least the local structures of $p(\boldsymbol{x})$.

It can be considered also that the SOM achieves a non-uniform quantization that transforms $\boldsymbol{x}$ to $\boldsymbol{w}_{\boldsymbol{x}}$ by minimizing the given metric. Nevertheless, thanks to the training phase (detailed below) the neurons $\boldsymbol{w}$ are located on the map according to their similarity. Then, when considering neurons $\boldsymbol{w}(i, j)$ located not *too far* from the winning neuron $\boldsymbol{w}_{\boldsymbol{x}}$, the distance in $\mathbb{R}^p$ between $\boldsymbol{x}$ and $\boldsymbol{w}(i, j)$ is not dramatically different from the one between $\boldsymbol{x}$ and $\boldsymbol{w}_{\boldsymbol{x}}$. That means that in the neighborhood of $\boldsymbol{w}_{\boldsymbol{x}}$ on the map (*i.e.* with closed location i and j), are located the wining neurons of the neighbors of $\boldsymbol{x}$ in $\mathbb{R}^p$. Hence, a class in the feature space $\mathbb{R}^p$ is projected into the map at the same area, remaining homogeneous. Moreover, whatever the initial shape of the class in the $\mathbb{R}^p$ feature space, the projected class is highly likely to be of isotropic shape in the map.

1) Training Phase: The learning phase may be thought of as a classification phase, such as a K-means classification algorithm. Neurons are first sampled (in $\mathbb{R}^p$) randomly and then, iteratively in a similar way as in the K-means algorithm, they are modified to fit a training sample $\mathcal{L} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L\}$. One of the main difference from the K-means algorithm is that the nodes which are close to the best-matching node in the map will learn from the same input $\boldsymbol{x}$ also.

While the initial values of the $\boldsymbol{w}$ may be set randomly, they will converge to a stable value at the end of the training process, by using (16):

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + h_{\boldsymbol{w},\boldsymbol{w}_{\boldsymbol{x}}}(t)\,(\boldsymbol{x} - \boldsymbol{w}(t))\,, \tag{16}$$

where t is the iteration index.

During one iteration of the training phase, every input $\boldsymbol{x}_\ell$, taken from the training set $\mathcal{L}$, is processed according to (16). $h_{\boldsymbol{w},\boldsymbol{w}_{\boldsymbol{x}}}(t)$ is called *neighborhood kernel*: it is a function defined over the lattice points of Kohonen's map, usually $h_{\boldsymbol{w},\boldsymbol{w}_{\boldsymbol{x}}}(t) = h(d(\boldsymbol{w}, \boldsymbol{w}_{\boldsymbol{x}}), t)$ where $d(\boldsymbol{w}, \boldsymbol{w}_{\boldsymbol{x}})$ stands for the distance between the location of $\boldsymbol{w}$ and $\boldsymbol{w}_{\boldsymbol{x}}$ on the map. While increasing $d(\boldsymbol{w}, \boldsymbol{w}_{\boldsymbol{x}})$, or increasing t, $h_{\boldsymbol{w},\boldsymbol{w}_{\boldsymbol{x}}}(t)$ decreases monotonically to 0. The average width and the form of $h_{\boldsymbol{w},\boldsymbol{w}_{\boldsymbol{x}}}(t)$, defines the "stiffness" of the "elastic surface" to be fitted to the data set. Let their index in the neighborhood

of $\boldsymbol{w}_{\boldsymbol{x}}$ be denoted by the set $\mathcal{N}_{\boldsymbol{w}_{\boldsymbol{x}}}(t)$ (see Fig. 1).

$$h_{\boldsymbol{w},\boldsymbol{w}_{\boldsymbol{x}}}(t) = \begin{cases} \alpha(t)\, \exp(-\frac{d(\boldsymbol{w},\boldsymbol{w}_{\boldsymbol{x}})}{2\sigma^2(t)}) & \text{if } \boldsymbol{w} \in \mathcal{N}_{\boldsymbol{w}_{\boldsymbol{x}}}(t), \\ 0 & \text{if } \boldsymbol{w} \notin \mathcal{N}_{\boldsymbol{w}_{\boldsymbol{x}}}(t). \end{cases} \quad (17)$$

The value of $\alpha(t)$ is then identified with a *learning-rate factor* ($0 < \alpha(t) < 1$). Both $\alpha(t)$ and the support of $\mathcal{N}_{\boldsymbol{w}_{\boldsymbol{x}}}(t)$ are usually decreasing monotonically in time (during the ordering process). $\sigma(t)$ is the width of the neighborhood that corresponds to the radius of the neighborhood of $\boldsymbol{w}_{\boldsymbol{x}}$ in $\mathcal{N}_{\boldsymbol{w}_{\boldsymbol{x}}}(t)$. In practice, $\alpha(t)$ and $\sigma(t)$ vanish with time. Typically, linearly decreasing functions are defined such as: $\alpha(t) = \alpha_0 \times \frac{T-t}{T}$ and $\sigma(t) = \sigma_0 \times \frac{T-t}{T}$, where T stands for the number of iterations.

2) Projection: Once the SOM has been trained, it acts as a similar way to as a set of clusters yielded by a K-means algorithms. Here, the index of each $\boldsymbol{w}$ is defined in 2D and each $\boldsymbol{w}$ located in the same area of the map has similar value in $\mathbb{R}^p$.

For each sample $\boldsymbol{x}$ to be processed, it is "projected" on the map by using eq. (15) to find its corresponding neuron $\boldsymbol{w}_{\boldsymbol{x}}$. The SOM may be considered to as a non-uniform quantization of the feature space [18]. This non-uniform quantization performed by Kohonen's map has the advantage to make the class definition on the map (*i.e.* through the quantization index) more isotropic than in $\mathbb{R}^p$. Then, the map may be considered to as an approximation in $\{1,\ldots,N\} \times \{1,\ldots,M\}$ of the initial manifold of $\mathbb{R}^p$, while preserving its topology.

B. Feature space for smart BBA

The proposed smart BBA intends to evaluate the mass of each class in 2^Θ or D^Θ according to the topology of the observed manifold. Then, two sets of data may be handled (see Fig. 2): on the first hand the initial observations $\boldsymbol{x}$ and class centers $\{C_1, C_2, \ldots, C_K\}$ in $\mathbb{R}^p$ and, on the other hand the so-called *winning neurons* $\boldsymbol{w}_{\boldsymbol{x}}$ and the projected class centers $\boldsymbol{w}_{C_k}$. It is worth noting that there is no link between the training of the classifier that defines $\{C_1, C_2, \ldots, C_K\}$ in $\mathbb{R}^p$ and the SOM that defines the set of neurons $w(i,j)$ in $\mathbb{R}^p$, $1 \leqslant i \leqslant M, 1 \leqslant j \leqslant N$, except that both are trained by using the same training samples (or a part of those).

$\boldsymbol{w}_{\boldsymbol{x}}$ is determined following eq. (15) and for $k \in \{1,\ldots,K\}$, $\boldsymbol{w}_{C_k}$ is determined in a similar way as stated in the following equation:

$$\boldsymbol{w}_{C_k} = \underset{\substack{w(i,j) \\ 1\leqslant i\leqslant M, 1\leqslant j\leqslant N}}{\arg\min} \|C_k - \boldsymbol{w}(i,j)\|. \quad (18)$$

Then, Kohonen's map can be used to build easily BBA and to balance between conjunction and disjunction when considering relative distance of an observation to the map. Moreover, the use of Kohonen's map simplifies the evaluation of the masses since operations on the maps require calculation on index only, while operations on the feature space (in $\mathbb{R}^p$) may be much more complex (when dealing with stochastic divergence for instance). So two kinds of distances will be considered and their related difference will induce uncertainty:

1) $d_{\mathbb{R}^p}(\cdot,\cdot)$ which is the distance in $\mathbb{R}^p$. It can be defined through the euclidean norm $\mathcal{L}^2(\mathbb{R}^p)$ but also through a spectral point of view such as the spectral angle mapper or the spectral information divergence [19]. It may also be based on the Kullback-Leibler divergence or the mutual information when dealing with Synthetic Aperture Radar (SAR) [20].
2) $d_{\text{map}}(\cdot,\cdot)$ which is the distance along Kohonen's map. It is mainly based on the euclidean norm and uses the index that locates the two vectors on the map: $d_{\text{map}}(\boldsymbol{w}_1, \boldsymbol{w}_2) = \sqrt{(n_1-n_2)^2 + (m_1-m_2)^2}$ if $\boldsymbol{w}_1$ (resp. $\boldsymbol{w}_2$) is located at position (n_1, m_1) (resp. (n_2, m_2)) on the map.

C. Mass function construction

This section details a method for building a BBA by using Kohonen's map and an initial classifier on $\mathbb{R}^p$.

1) Mass of simple hypotheses: The definition of masses of focal elements could be based on the distance on the feature space. Nevertheless, an appropriated definition should take into account the variance of the classes to weight each of them, as it is the case in a likelihood point of view. This weighting is already performed by the projection onto Kohonen's map so that, the mass of focal class is defined as:

$$\begin{cases} m(\boldsymbol{x} \in \theta_k) & \sim 1 \qquad \text{if } \boldsymbol{w}_{\boldsymbol{x}}{=}\boldsymbol{w}_{\mathcal{C}_k} \\ m(\boldsymbol{x} \in \theta_k) & \sim \dfrac{d_{\text{map}}(\boldsymbol{w}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_k})^{-1}}{\sum_{\ell=1}^{K} d_{\text{map}}(\boldsymbol{w}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_\ell})^{-1}} \quad \text{otherwise} \end{cases} \quad (19)$$

where $k = 1, 2, \ldots, K$, $\boldsymbol{w}_{\mathcal{C}_k}$ is the projected class, $\boldsymbol{w}_{\boldsymbol{x}}$ is the winning neurons.

According to eq. (19), we consider that the more the distance $d_{\text{map}}(\boldsymbol{w}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_k})$ (relatively to the other distances between $\boldsymbol{x}$ and $\mathcal{C}_\ell$ on the map) the less the mass $m(\boldsymbol{x} \in \theta_k)$.

2) Mass of the full ignorance: From the feature space, we consider that the mass evaluation of an observation falls into ignorance if its distance to the map is much more important that the distance of its related class center to the map. Then, it can be expressed as follows:

$$m(\boldsymbol{x} \in \Theta) \sim 1 - \min\left(\frac{d_{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{w}_{\boldsymbol{x}})}{d_{\mathbb{R}^p}(\mathcal{C}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}})}, \frac{d_{\mathbb{R}^p}(\mathcal{C}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}})}{d_{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{w}_{\boldsymbol{x}})} \right) \quad (20)$$

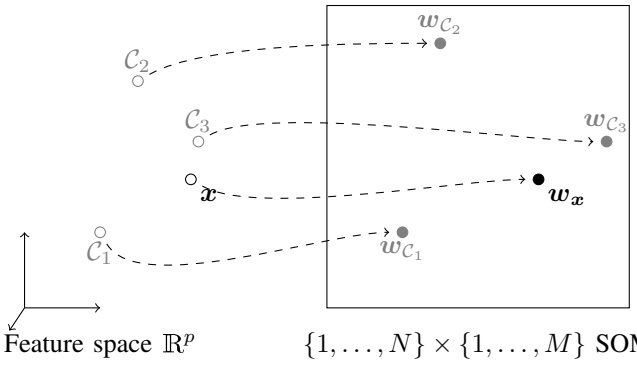


Fig. 2. Observations in the feature space and their projections into Kohonen's map. Note that the neurons $\boldsymbol{w}_{\boldsymbol{x}}$ and $\boldsymbol{w}_{C_k}$ can be located on the map through their location index (n,m) or in $\mathbb{R}^p$ with their p component value.

where $\mathcal{C}_{\boldsymbol{x}}$ is the class center of $\boldsymbol{x}$, $\boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}}$ is its projection on the map.

3) Mass of the conjunction between two classes: In the set D^{Θ}, the conjunction between two classes may be defined into the feature space as the space in-between the two classes. But, one has to account for the variance of each classes that increases the complexity of this measure. Once again, it is much more convenient to define the $\theta_k \cap \theta_\ell$ mass directly into Kohonen's map, as:

$$m(\boldsymbol{x} \in \theta_k \cap \theta_\ell) \sim e^{-\gamma(z-1)^2} \tag{21}$$

with

$$z = d_{\text{map}}(\boldsymbol{w}_{\boldsymbol{x}}, \frac{\boldsymbol{w}_{\mathcal{C}_k} + \boldsymbol{w}_{\mathcal{C}_\ell}}{2}) \qquad 0 < k, \ell \leqslant K, \ell \neq k$$

Eq. (21) stipulates that the value of $m(\boldsymbol{x} \in \theta_k \cap \theta_\ell)$ becomes maximal when $\boldsymbol{x}$ reaches the middle of $[\boldsymbol{w}_{\mathcal{C}_k}, \boldsymbol{w}_{\mathcal{C}_\ell}]$ segment. Eq. (21) yields a value of $m(\boldsymbol{x} \in \theta_k \cap \theta_\ell)$ closed to 1 in the middle. Moreover, $m(\boldsymbol{x} \in \theta_k \cap \theta_\ell)$ vanishes when $\boldsymbol{x}$ is far away from the $[\boldsymbol{w}_{\mathcal{C}_\ell}, \boldsymbol{w}_{\mathcal{C}_k}]$ segment. The γ parameter tunes this vanishing behavior. For example, if we want eq. (21) be over $\frac{1}{2}$ between the 1st and the 3rd quartile of $[\boldsymbol{w}_{\mathcal{C}_k}, \boldsymbol{w}_{\mathcal{C}_\ell}]$ segment, then γ should be equal to $2\sqrt{2}$. For a smaller domain around the median of $[\boldsymbol{w}_{\mathcal{C}_k}, \boldsymbol{w}_{\mathcal{C}_\ell}]$ segment, γ should be greater (see Fig. 3).

This conjunctive mass estimation does not apply in the classical Dempster-Shafer framework (*i.e.* when working in 2^{Θ} only assuming Shafer's model of the frame Θ).

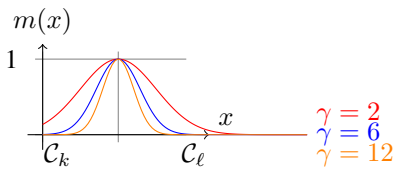


Fig. 3. Behavior or $m(\boldsymbol{x} \in \theta_k \cap \theta_\ell)$ with γ, according to eq. (21).

4) Mass of disjunction between two classes: The ignorance in the decision making between two classes $\mathcal{C}_k$ and $\mathcal{C}_\ell$ may be considered as the dual of eq. (21), but here by considering distances in the feature space. When a sample $\boldsymbol{x}$ is not *too far* from class $\mathcal{C}_k$ or $\mathcal{C}_\ell$, it is not *too difficult* to decide if it has too be associated to the class k or ℓ. But if $\boldsymbol{x}$ is *far* from $\mathcal{C}_k$ and $\mathcal{C}_\ell$, it comes the disjunction as related in Fig. 4. That corresponds to a context where the distances between $\boldsymbol{x}$ and the classes are of the same scale: $d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k) \approx d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)$. But such criteria is not enough since it includes also the case where $\boldsymbol{x}$ is located in-between $\mathcal{C}_k$ and $\mathcal{C}_\ell$. So it has to be weighted by the distance between the two classes $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)$. If $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell) \ll d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k)$ and $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell) \ll d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)$, $\boldsymbol{x}$ falls in the disjunctive case since $\boldsymbol{x}$ is considered *far* to $\mathcal{C}_k$ and $\mathcal{C}_\ell$. Then, the criteria defined in eq. (22) is based on the ratio between $d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)$ and $d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)$.

Then, the mass of the disjunction $\theta_k \cup \theta_\ell$ is modeled by:

$$m(\boldsymbol{x} \in \theta_k \cup \theta_\ell) \sim 1 - \tanh(\beta z) \tag{22}$$

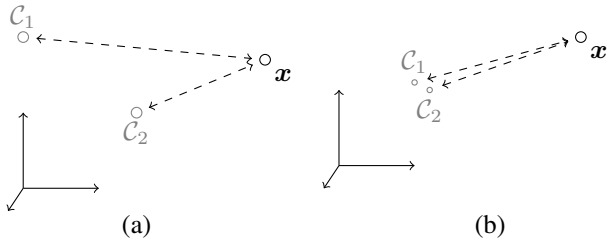


Fig. 4. Disjunction between two class: (a) non ambiguous case, (b) ambiguous case.

with

$$z = \frac{d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)}{d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)} \qquad 0 < k, \ell \leqslant K, k \neq \ell.$$

Here, the β parameter stands for the level of ambiguity. When $\boldsymbol{x}$ is close, in $\mathbb{R}^p$, to the segment $[\mathcal{C}_k, \mathcal{C}_\ell]$, $d(\mathcal{C}_k, \mathcal{C}_\ell) \simeq d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)$ so that z is close to 1, and $m(\boldsymbol{x} \in \theta_k \cup \theta_\ell)$ has to vanish. Then, the areas where eq. (22) vanishes are shown on curves of Fig. 5. The more the β, the less the ambiguous mass.

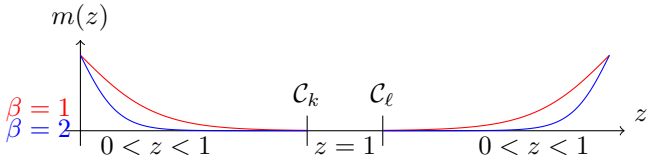


Fig. 5. Shape of eq. (22) for some value of β.

5) Conjunction and disjunction for more than 2 classes: This construction that takes into consideration the ratio of distance between 2 classes or the distance to the middle of 2 classes can be extended to more than 2 classes. For instance, eq. (21) can be based on the centroid of more than 2 class. Eq. (22) can be generalized by the composition of one against one class from a set of K classes, divided by the sum of distance of $\boldsymbol{x}$ to each of the K class centers. Nevertheless, this method of construction has not been deeper investigated since those compositions should not have significative impact on the fusion or the classification results.

6) Normalized BBA: the complete BBA has to respect eq. (1) constraint so that is it necessary to apply a normalization step to the unnormalized BBA obtained by separately calculates the belief masses on simple and compound hypotheses, presented in previous steps 1–4.

7) Determination of parameters β and γ: The determination of the parameters β and γ can be found automatically by minimizing the following constraints, defined in [21], [5]:

$$E = \sum_{i=1}^{\text{num Samples}} \sum_{n=1}^{N} (BetP(\boldsymbol{x}_i \in \theta_n) - \Upsilon(\boldsymbol{x}_i \in \theta_n))^2$$

where $BetP(\boldsymbol{x}_i \in \theta_n)$ stands for the pignistic probability of x_i (vector to classify) according to the simple hypothesis θ_n, and $\Upsilon(\boldsymbol{x}_i \in \theta_n)$ is a function that is equal to 1 if the sample $\boldsymbol{x}_i$ does belong to the simple hypothesis θ_n (as stated *a priori*

from the learning base), and 0 otherwise.

IV. Simple simulation

This section presents a simulation dedicated to a simple 4-class problem. Although SOM is more appropriated to be used to perform a non linear projection from $\mathbb{R}^p$ to $\{1,\dots,N\}\times\{1,\dots,M\}$ with $p>2$, this naive case of study has been defined in $\mathbb{R}^2$ for a better visualisation. Fig. 6 shows, with black crosses, a dataset in $\mathbb{R}^2$ that is decomposed into 4 clusters. Each of those clusters have a gaussian shape with different covariance matrices.

The classifiation yielded by a K-means gives 4 clusters $\mathcal{C}_1$ to $\mathcal{C}_4$ which appear with green bullets in Fig. 6. Their locations are approximatively: $\mathcal{C}_1$: $(0.18, 0.18)$, $\mathcal{C}_2$: $(0.6, 0.18)$, $\mathcal{C}_3$: $(0.25, 0.4)$ and $\mathcal{C}_4$: $(0.45, 0.5)$ which corresponds to the center of each gaussian sampling.

When performing a Kohonen's map of size 8×8, it yields the map characterized by the red bullets in Fig. 6. As drawn in the $\mathbb{R}^2$ feature space, the map is seens dramatically deformed according to the density of data samples. The more the density of samples (in black crosses), the more the density of the neurons (in red bullets), which is a characteristic of a non-uniform quantization. The location of the red bullets corresponds of the value of the weight of the neurons in $\mathbb{R}^2$.

Then, when a sample (black cross) is *projected* into the map, it is associated to its winning neuron according to eq. (15), *i.e.* associated to the closest red bullet according to the euclidean distance in $\mathbb{R}^2$. Fig. 7-(a) shows the same figure as Fig. 6 highlighting some areas. The ellipses in blue highlight the areas between the different clusters while

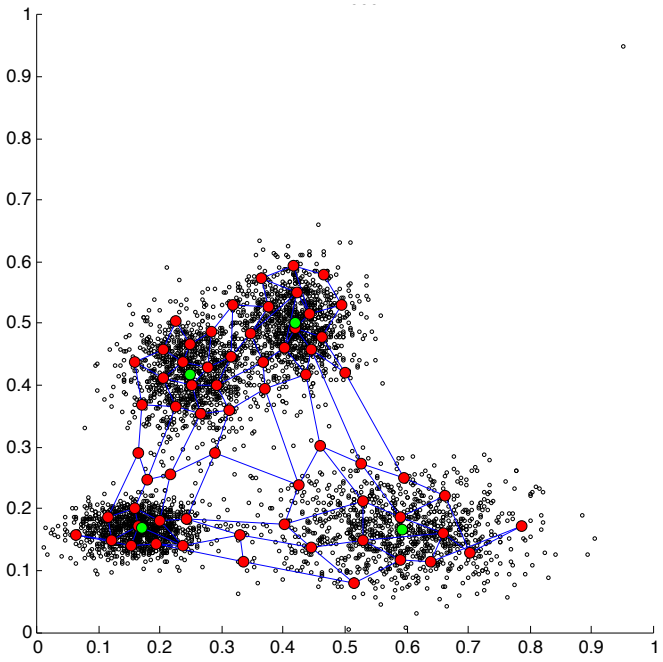
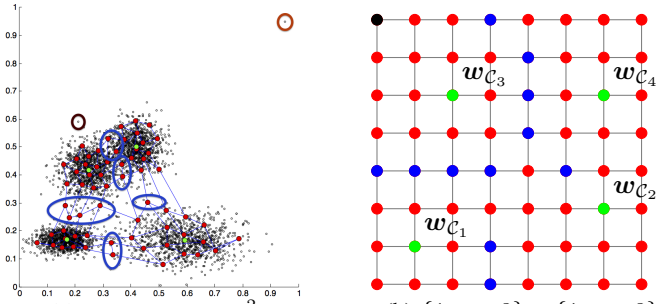


Fig. 6. Simple simulation of a 4-class manifold with an outlier. Samples of the data set in $\mathbb{R}^2$ are shown with black crosses, the red bullets characterize the locations of the neurons of Kohonen's map. Blue lines show the SOM projected in the feature space. The 4 green bullets characterise the K-means class centers.



(a) Feature space $\mathbb{R}^2$ (b) $\{1,\dots,8\}\times\{1,\dots,8\}$ SOM

Fig. 7. Simple simulation of Fig. 6 with its equivalent in the SOM geometry. Green bullets on the map correspond to the winning neurons $w_{\mathcal{C}_k}$ of the class centers $\mathcal{C}_k$, blue bullets, on the map in (b), correspond to the location of the neurons that are located in between classes in $\mathbb{R}^2$. The black neuron at the top left of the map corresponds to the winning neuron of the sample rounded with a brown ellipse at location $(0.2, 0.6)$.

the red ellipse at the top right of the figure points out an outlier. On Fig. 7-(b) is shown Kohonen's map into its natural geometry in $\{1,\dots,N\}\times\{1,\dots,M\}$: the distance between neurons corresponds to the distance along the edges of the map, *i.e.* considering the indexes. The green bullets in this map shows the winning neurons $w_{\mathcal{C}_k}$ of the class centers $\mathcal{C}_k$. The neurons shown in blue correspond to the neurons rounded by the ellipses in blue in Fig. 7-(a). Those neurons are located between classes in $\mathbb{R}^2$ and also between the corresponding class centers $w_{\mathcal{C}_k}$ and $w_{\mathcal{C}_\ell}$. This point illustrates the topological preservation of Kohonen's map.

Let us focus on the sample at location $(0.2, 0.6)$ which is rounded with a brown ellipse. A first look at Fig. 7-(a) points out that this sample is located very near class $\mathcal{C}_3$ but a little bit outside the main concentration of the dataset. The winning neuron associated to this sample is drawn in black in Fig. 7-(b), at the top left of the map (with index location $(1, 8)$). It is clear that this neuons is closed to $w_{\mathcal{C}_3}$ and far from the winning neurons of the other classes. Then, the second maximum of the BBA reach the mass $m(\boldsymbol{x}\in\theta_3\cap\theta_4)$ (with a value of 0.1135) and the third is devoted to $m(\boldsymbol{x}\in\Theta)$ (at 0.1005). Considering eq. (22) Following eq. (20), it appears that $d_{\mathbb{R}^p}(\boldsymbol{x},\boldsymbol{w}_{\boldsymbol{x}})$ is of significant value un comparison of $d_{\mathbb{R}^p}(\mathcal{C}_3,\boldsymbol{w}_{\mathcal{C}_3})$ so that $m(\boldsymbol{x}\in\Theta)$ has also a significant value.

Let us focus on the outlier located at $(0.95, 0.95)$ at the top right of Fig. 6. This sample is located at the top right of Fig. 7-(a). It is far from the rest of the data set and also far from Kohonen's map. Its winning neuron is located at position $(8, 8)$ (*i.e.* at the top right of the map in Fig. 7-(b)). Since this neuron is closed to $w_{\mathcal{C}_4}$ it is expected that the mass $m(\boldsymbol{x}\in\theta_4)$ be significant. It is the case with a value of 0.1798. Nevertheless, the maximum value of the BBA is reached with $m(\boldsymbol{x}\in\Theta)$ with 0.2660 which underlines the outlier behavior of this sample. The resulting BBA is very informative because the rest of the masses vanish below 0.09.

Let us focus now in a sample located at position $(0.2, 0.3)$ in Fig. 7-(a). This point is in the middle of two classes $\mathcal{C}_1$ and $\mathcal{C}_3$. A little bit closer to class $\mathcal{C}_1$. Its winning neuon falls in the blue dots of Fig. 7-(b) at location $(3, 4)$. Then the mass of $m(\boldsymbol{x}\in\theta_1\cap\theta_3)$ traps a significant value as high as 0.1313.

Nevertheless, the second highest value of this BBA is reached by $m(\boldsymbol{x} \in \theta_1 \cap \theta_4)$ with 0.1102. The fact is that, considering the location of the winning neuron in Kohonen's map, it is near to the middle of $w_{\mathcal{C}_1}$, $w_{\mathcal{C}_2}$ and $w_{\mathcal{C}_4}$. The third maximum falls to $m(\boldsymbol{x} \in \theta_2 \cup \theta_4)$ (value 0.0831).

This simple example shows that the aim of this BBA modeling technique that induces simple consideration on the distance from samples to clusters in the feature space $\mathbb{R}^p$ and in Kohonen's map $\{1, \ldots, N\} \times \{1, \ldots, M\}$.

V. Experiments on benchmark dataset

In order to highlight some advantages and possible drawbacks of the proposed SOM-based BBA modeling, the performance of the SOM-based BBA is compared to EVCLUS and ECM ones by using dataset provided by the University of California - Irvine (UCI) Machine Learning Repository[3]. 7 data sets out of 270 have been taken into consideration with various amount of features (that corresponds to the feature space dimension $\mathbb{R}^p$) and number of classes (from 2 to 7) as detailed in Table I.

In this section, experimental results are based on the classical Dempster-Shafer framework (i.e. we work with 2^Θ only). Indeed, ECM, EVCLUS are only working in this framework.

It is worth noting that the Matlab programs of ECM and EVCLUS have been downloaded from the official webpage page of Thierry Denœux for those experiments[4]. Most of the internal parameters have been let ti their default value. The distance δ to the empty set has been changed to 100 in ECM and the regularization parameter has been changed to 0.5 in EVCLUS. The number of clusters in ECM and EVCLUS has been fixed accroding to Table I, depending on the dataset.

Kohonen's map has been trained with the following parameters: a size of 20×20 neurons (except for Seeds and Wine a size of 10×10 neurons), trained with 200 iterations. An initial neighborhood size $\mathcal{N}_{\boldsymbol{w}}(t_0)$ of 10 neurons and a learning rate $\alpha(t_0)$ of 0.9. These values were carefully selected in order to guarantee convergence of the map with appropriate number of neurons to well balance the tradeoff between quantization error and manifold approximation, so as to improve results.

[3]The dataset is available at http://archive.ics.uci.edu/ml

[4]Thierry Denœux's webpage is available at https://www.hds.utc.fr/~tdenoeux/dokuwiki/en/software.

TABLE I
Characteristics of the UCI datasets used for comparison.

Dataset	Features	classes	samples
Banknote authentication	4	2	1372
Pima Indians Diabetes	8	2	768
Seeds	7	3	210
Wine	13	3	170
Statlog (Landsat Satellite)	36	6	6435
Statlog (Image Segmentation)	19	7	2130
Synthetic control chart time series	60	6	600

The quantization error through the Root Mean Squared Error (RMSE) is used here as criterion to evaluate the quality of Kohonen convergence:

$$\text{EQM} = \left(\frac{1}{\text{num Sample}} \sum_{\ell=1}^{\text{num Sample}} \|\boldsymbol{x}_\ell - \boldsymbol{w}_{\boldsymbol{x}_\ell}\|^2 \right)^{\frac{1}{2}}.$$

In this section, the values of the parameter β has been selected based on the results shown in Table II. In this experiment, $\beta = 2$ yields the best classifications results. Also, it can be noticed that the proposed method is not so sensitive to the value of β.

TABLE II
Classification results of SOM-based BBA in 2^Θ for different value of β.

Dataset	$\beta = 1$	$\beta = 2$	$\beta = 6$
Seeds	87.6190%	**90.9524%**	89.0476 %
Wine	71.1765%	**73.5294 %**	71.7647%

It appears that the SOM-based BBA yields most of the time the highest classifications results. Those best results are shown in bold of Table III, where the first line corresponds to the number of correctly classified samples, the second line corresponds to the proportion of samples correctly classified the last line shows the computation time. When ECM appears better, SOM-based approach is close to the best accuracy (73.52 % versus 74.11 % for the benefit of ECM with the Wine database, and 69.24 % versus 69.62 % with the Statlog Landsat satellite images database). EVCLUS is always below. It seems that the performance ranking between ECM and SOM-based

TABLE III
Classification results in 2^Θ of EVCLUS, ECM and SOM-based BBA with decision by the maximum of pignistic probability.

Dataset	EVCLUS	ECM	SOM-based
Banknote authentication	843 61.44 % 1172.2sec	848 61.80 % 3.4sec	1090 **79.44 %** 8.6sec
Pima Indians Diabetes	475 61.84 % 181.7sec	506 65.88 % 3.2sec	549 **71.48 %** 6.7sec
Seeds	157 74.76 % 34.3sec	189 90.0 % 0.3sec	191 **90.95 %** 5.8sec
Wine	103 60.58 % 6.7sec	126 **74.11 %** 0.9sec	125 73.52 % 5.9sec
Statlog (Landsat Satellite)	3027 47.03 % 5857sec	4480 **69.62 %** 480sec	4456 69.24 % 163sec
Statlog (Image Segmentation)	895 42.01 % 3657sec	1282 55.49 % 161sec	1431 **67.18 %** 84sec
Synthetic control chart time series	384 64.0 % 370sec	453 72.5 % 6.9sec	501 **83.5 %** 8.0sec

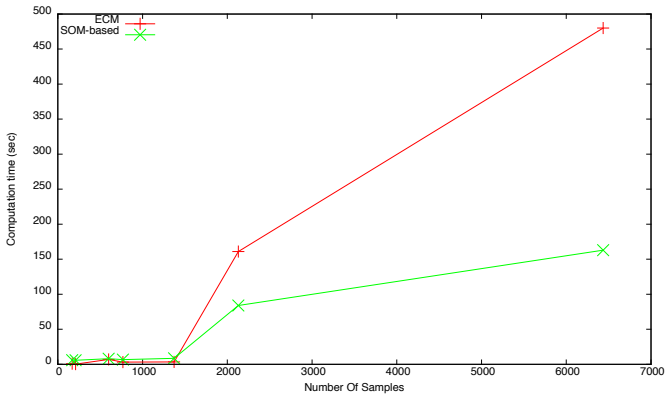


Fig. 8. Computation time depending on the feature space dimension. SOM-based approach is more appropriated for processing large amount of data than ECM.

BBA does not depend on the feature space dimension nor the number of classes since the Wine and Statlog Landsat satellite image data bases are very different to each other. Since the SOM-based approach considers a projected feature space of dimension 2, it may induce on those cases a too coarse approximation of the manifold in comparison to ECM. Nevertheless, it is worth noting that the benefit in using a SOM-based approach for BBA is related to the number of samples to be handled [22]. Fig. 8 shows that the more the number of sample the fastest the SOM-based approach in comparison to the ECM while yielding the same level of accuracy. Then the SOM-based approach appears to be a valuable alternative to handle large data set such as real images for classification purpose.

VI. Experiments on a real satellite image

The proposed methodology is now applied on a SPOT image ($1318 \times 2359 = 3$ Mega pixels) taken in 2000 for classification purpose. From the variety of objects constituting this image, five clusters may be distinguished: Covered Fields (CF) light-red area, Bare soil (BS) red area, Wooded Area (WA) dark-red area, Water or Wet area (WWA) green area and Bare Soil and Wet Area (BSWA) bright-green area (see Fig. 9). Those five classes will constitute our frame of discernment $\Theta = \{CF, BS, WA, WWA, BSWA\}$. This 3-band multi-spectral image represents a single source of information in $\mathbb{R}^3$, so that there is no fusion process within the components of each pixel for BBA (except DC which uses eq. (9) to perform a fusion rule class by class). In this experiment, DC, ECM and the SOM-based methods are tested.

Kohonen's map has been trained with the same parameters as in the previous section.

A. The classification results in 2^Θ

In order to generate mass function on the disjunction of hypotheses in DC, Dempster's combination rule given by eq. (9) has been replaced by the disjunctive rule given by eq. (2). Fig. 10 shows the classification of the original image with DC approach and the proposed approach by using the criterion of the maximum of pignistic probability for decision-making on simple hypotheses (classes). Fig. 11 shows the classification results all over simple classes and all disjunctions of classes. The performance of the classifiers is shown through the confusion matrix form in Table V. The test has been done over 16692 pixels where 3273 represent Covered Fields, 2273 Wooded Area, 3013 Bare Soil, 6005 Water or Wet area and 10 Bare Soil and Wet Area. The legend (colors of decision classes in the images classification), is given in Table IV.

TABLE IV
DST legend used on classification results of Fig. 11.

WWA	BSWA ∪ BS	BSWA ∪ CF
BSWA	BSWA ∪ WA	BS ∪ WWA
BS	BS ∪ WA	BS ∪ CF
WA	WWA ∪ WA	WWA ∪ CF
CF	BSWA ∪ WWA	WA ∪ CF

As Table V shows, the SOM-based approach presents promising results. Indeed by comparing our approach to the DC approach, it can be noticed that class detection has been improved. In Fig. 10-(a), the river is well discriminated in comparison to other classes while in Fig. 10-(b) a great conflict appears when those classes WW and BSW have to be discriminated. Fig. 11 demonstrates that our approach reduces the number of decision class (8 classes), whereas DC approach yields multiple classes. For example the whole river is almost attributed to a single class; this reflects more what we have in the reality, while with the other approaches the river is classified into various class.

After having exploring the performance of the SOM-based approach, this part focuses on the ability of the SOM-based approach to deal with a large amount of multi-variate data. To evaluate this, the unsupervised clustering method ECM has been used for its simplicity in generating the BBA in the case of exploratory data analysis. This algorithm requires a great amount of computing time for processing the large images. Here, a crop of the original image (300 by 220 pixels)

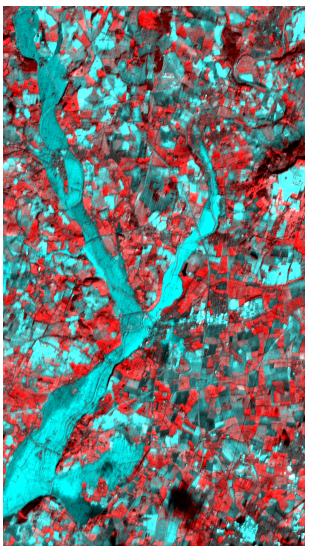


Fig. 9. A false color composite of the SPOT image. ©CNES.

has been processed so that the computation time remains acceptable. The classification results (see Fig. 12) show that the SOM-based method gives higher performances than the ECM algorithm, while remarkably reducing the computational cost. Indeed, SOM-based method has a linear computational complexity depending to the number of classes for each new calculated BBA. These results prove that the proposed approach provides a very significant advantage in the case of processing large images.

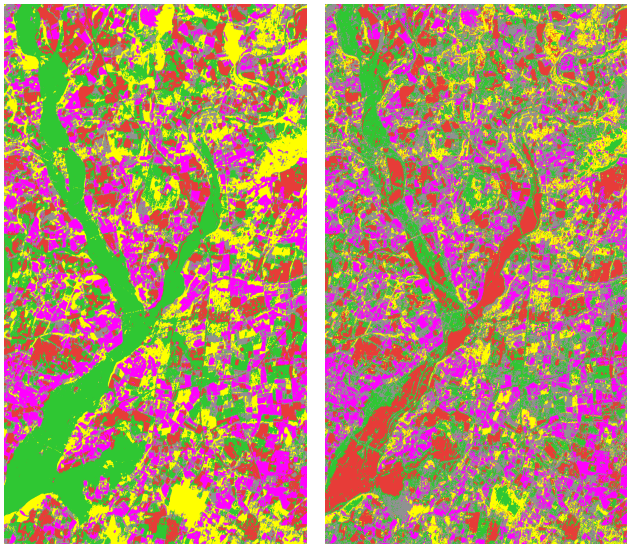
All the algorithms in the experiments were coded in MATLAB^{TM} without specific optimization and run on a machine with 3.4 GHz Intel Core i7-3770M processor and 8 GB memory running the Windows 7 Server operating system. The execution times for these algorithms are: 20 minutes and 12 seconds for the SOM-based BBA shown in Fig. 12-(a), and 2 days and 6 hours and 45 seconds for the ECM algorithm shown in Fig. 12-(b). It corresponds to an increase in computation speed of 150!

The computation of the complete scene at Fig. 10-(a) took 4 hours and 21 minutes and 36 seconds.

B. The classification results in D^{Θ}

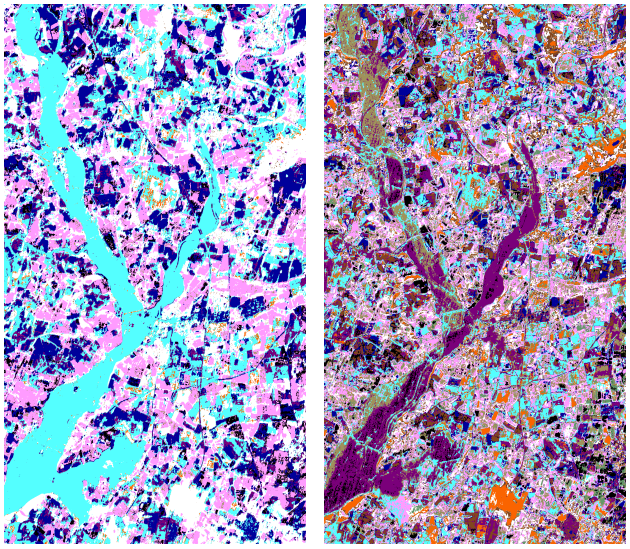
In this experiment, the result of DC is given by replacing Dempster's combination rule given by eq. (9) by the conjunctive rule given by eq. (3). Fig. 13 shows the classification of the original image by using maximum of generalized pignistic probability over all simples classes and all conjunctions of classes. The performance of the classifiers is shown in the confusion matrix form in Table VII. Table VI represents the colors assigned to each conjunctions of classes in classification. The colors assigned to simple classes and to disjunctions of classes are the same as those defined in Table IV.

As shown in Table VII, the generation of the masses on the conjunctions of hypotheses has degraded remarkably the DC result. This is due to the conflicting nature of the conjunctive rule when unreliable sources are combined. The SOM-based approach can overcome this problem by calculating the masses of conjunctions from Kohonen's map.



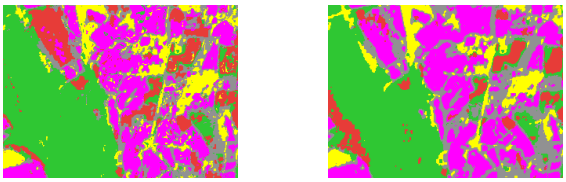
(a) (b)

Fig. 10. Classification results in 2^{Θ} with decision by maximum of pignistic probability over all simples hypotheses: (a) SOM-based BBA. (b) DC results.



(a) (b)

Fig. 11. Classification results in 2^{Θ} with decision by maximum of pignistic probability over all simples hypotheses and all disjunctions of hypotheses: (a) SOM-based BBA. (b) DC results.



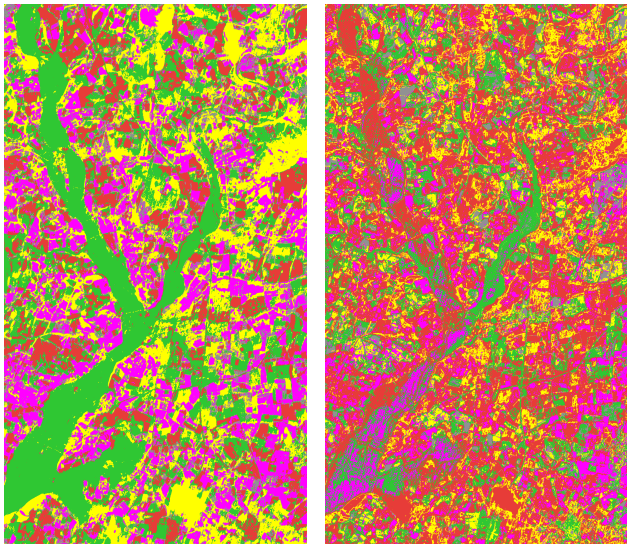
(a) (b)

Fig. 12. Classification results in 2^{Θ} with decision by maximum of pignistic probability: (a) SOM-based approach. (b) ECM.

Fig. 14 shows the classification of the original image by using maximum of generalized pignistic probability over all simples classes, all conjunctions of classes and all disjunctions of classes. As seen in DST-based experiment, it appears that the SOM-based approach yields promising results with a very reasonable computation time in such situations even with large number of classes.

TABLE V
COMPARATIVE CLASSIFICATION RESULTS IN 2^{Θ}. COMPARISON BETWEEN DC APPROACH AND THE PROPOSED SOM-BASED APPROACH.

		BSWA	BS	WWA	CF	WA
BSWA	DC	2102	0	0	0	26
	SOM	1631	282	215	0	0
BS	DC	87	2106	397	3	420
	SOM	206	2423	16	93	215
WWA	DC	2393	326	3092	194	0
	SOM	0	45	5359	601	0
CF	DC	180	206	819	1068	0
	SOM	0	103	53	2117	0
WA	DC	4	614	155	45	2455
	SOM	165	86	3	1	3018



(a) (b)

Fig. 13. Classification results in D^{Θ} with decision by maximum of generalized pignistic probability over all simples hypotheses and all conjunctions of hypotheses: (a) SOM-based approach. (b) DC results.

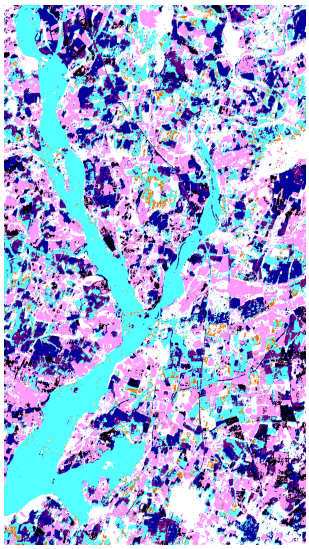


Fig. 14. Credal classification results in D^{Θ} through the SOM-based approach: maximum of generalized pignistic probability.

VII. Conclusion

The interest of evidence theory came from its ability to deal with uncertain and paradox data through the mass functions. Nevertheless, to the best of our knowledge, rare are the estimating mass functions approaches that consider the belief masses on compound hypotheses directly. In this research work, a new method for mass function construction through Kohonen's map has been proposed, and some experiments of the proposed method has been dedicated to image classification. The comparison with state-of-the art UCI database showed the accuracy of the SOM-based approach and its capability to deal with large amount of data. A further advantage can be added which is the possibility to perform the assignment of belief masses on the conjunctive and disjunctive hypotheses directly.

In this study, we focus on the application of the proposed Kohonen's map based BBA on SPOT images only, which is based on a quadratic distance evaluation. The extension to the problem of SAR image classification brings specific problems that have to be tackled. Once resolved, the credal fusion of optical and SAR remote sensing images for joint classification will be investigated.

TABLE VI
Legend used for classification D^{Θ} shown in Fig. 13

BSWA ∩ BS	BSWA ∩ CF
BSWA ∩ WA	BSA ∩ WWA
BSA ∩ WA	BSA ∩ CF
WWA ∩ WA	WWA ∩ CF
BSWA ∩ WWA	WA ∩ CF

References

[1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.

[2] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[3] F. Smarandache and J. Dezert, *Advances and Applications of DSmT for Information Fusion (Collected works), second volume: Collected Works*. Infinite Study, 2006.

[4] P. Smets, "Belief functions: The disjunctive rule of combination and the generalized bayesian theorem," in *Classic Works of the Dempster-Shafer Theory of Belief Functions*, ser. Studies in Fuzziness and Soft Computing, R. Yager and L. Liu, Eds., vol. 219. Springer Berlin Heidelberg, 2008, pp. 633–664.

[5] T. Denœux and L. M. Zouhal, "An Evidence-Theoric k-NN Rule with Parameter Optimization," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 28, no. 2, pp. 263–271, May. 1998.

[6] J. Schubert, "On nonspecific evidence," *International Journal of Intelligent Systems*, vol. 8, no. 6, pp. 711–725, 1993.

[7] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.

[8] F. Smarandache and J. Dezert, "Information fusion based on new proportional conflict redistribution rules," in *8th International Conference on Information Fusion*, vol. 2. Philadelphia, USA, Jul. 2005, pp. 1–8.

[9] M. C. Florea, J. Dezert, P. Valin, F. Smarandache, and A.-L. Jousselme, "Adaptative combination rule and proportional conflict redistribution rule for information fusion," in *COGnitive systems with Interactive Sensors*. Paris, France, mar. 2006.

[10] F. Smarandache and J. Dezert, *Advances and Applications of DSmT for Information Fusion*, A. R. Press, Ed., Rehoboth, 2004.

TABLE VII
Comparative classification results in D^{Θ}. Comparison between DC approach and the proposed SOM-based approach.

		BSWA	BS	WWA	CF	WA
BSWA	DC	0	1053	317	382	376
	SOM	1913	0	215	0	0
BS	DC	2056	90	86	6	255
	SOM	2080	1507	368	313	545
WWA	DC	2249	2	2081	24	1649
	SOM	0	0	5404	601	0
CF	DC	1536	0	0	737	0
	SOM	0	1	2272	0	0
WA	DC	1733	1	151	45	1334
	SOM	165	29	8	3	30068

[11] P. Smets, "Constructing the pignistic probability function in a context of uncertainty." in *UAI*, vol. 89, 1989, pp. 29–40.
[12] M.-H. Masson and T. Denœux, "ECM: An evidential version of the fuzzy C-means algorithm," *Pattern Recognition*, vol. 41, no. 4, pp. 1384–1397, apr. 2008.
[13] T. Denœux and M.-H. Masson, "EVCLUS: EVidential CLUSstering of proximity data," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 1, pp. 95–109, 2004.
[14] I. Borg and P. Groenen, *Modern multidimensional scaling*, Springer, Ed., New-York, 1997.
[15] A. Appriou, "Discrimination multisignal par la théorie de l'évidence," *Décision et Reconnaissance des formes en signal, Hermes Science Publication*, pp. 219–258, 2002.
[16] A. Dromigny-Badin, S. Rossato, and Y. Zhu, "Fusion de données radioscopiques et ultrasonores via la théorie de l'évidence," *Traitement du Signal*, vol. 14, no. 5, pp. 499–510, 1997.
[17] M. Kraaijveld, J. Mao, A. K. Jain *et al.*, "A nonlinear projection method based on kohonen's topology preserving maps," *Neural Networks, IEEE Transactions on*, vol. 6, no. 3, pp. 548–559, may. 1995.
[18] A. Czihò, B. Solaiman, G. Cazuguel, C. Roux, and I. Loványi, "Kohonen's self organizing feature maps with variable learning rate. application to image compression," in 3^{rd} *international workshop on Image and Signal Processing*, Santorini, Greece, 1997, pp. 11–14.
[19] C. Chang, "An information theoretic-based measure for spectral similarity and discriminability," *Information Theory, IEEE Transactions on*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.
[20] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal sar images and its extension to multiscale change analysis," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 5, pp. 1432–1445, 2007.
[21] T. Denœux, "A k-Nearest Neighbor classification rule based on Dempster-Shafer theory," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 25, no. 5, pp. 804–813, May. 1995.
[22] I. Hammami, J. Dezert, G. Mercier, and A. Hamouda, "On the estimation of mass functions using self organizing maps," in *Belief'2014*, sep. 2014, pp. 275–283.